Received: 28 January 2020 | Revised: 15 May 2020 | Accepted: 15 June 2020
DOI: 10.1002/sys.21550

REGULAR PAPER

WILEY

Assaying the importance of system complexity for the systems engineering community

Matthew W. Potts[1,*] | Pia A. Sartor[1,*] | Angus Johnson[2,*] | Seth Bullock[3,*]

[1] Aerospace Department, The University of Bristol, United Kingdom

[2] Thales Research, Technology and Innovation, Reading, Berkshire, England

[3] Department of Computer Science, The University of Bristol, United Kingdom

Correspondence
Matthew W. Potts, Aerospace Department, The University of Bristol, Bristol, Avonmouth, BS81TB, United Kingdom.
Email: matt.potts@bristol.ac.uk

*Equally contributing authors.

Funding information
Engineering and Physical Sciences Research Council, Grant/Award Number: 16000139; Industrial Cooperative Awards in Science & Technology

Abstract

How should organizations approach the evaluation of system complexity at the early stages of system design in order to inform decision making? Since system complexity can be understood and approached in several different ways, such evaluation is challenging. In this study, we define the term "system complexity factors" to refer to a range of different aspects of system complexity that may contribute differentially to systems engineering outcomes. Views on the absolute and relative importance of these factors for early–life cycle system evaluation are collected and analyzed using a qualitative questionnaire of International Council on Systems Engineers (INCOSE) members ($n = 55$). We identified and described the following trends in the data: there is little between-participant agreement on the relative importance of system complexity factors, even for participants with a shared background and role; participants tend to be internally consistent in their ratings of the relative importance of system complexity factors. Given the lack of alignment on the relative importance of system complexity factors, we argue that successful evaluation of system complexity can be better ensured by explicit determination and discussion of the (possibly implicit) perspective(s) on system complexity that are being taken.

KEYWORDS
complexity science, defense & security, government, system architecture

1 | INTRODUCTION

Organizations are increasingly having to engineer complex systems to meet the needs of a connected world.[1] There are several challenges inherent in engineering novel complex systems; such systems are generally made up of a large number of diverse, interdependent subsystems and components, interconnected via nonlinear relationships, leading to difficulties in predicting overall system performance.[2–8] System complexity has been shown to negatively affect system delivery project outcomes.[2] Therefore, organizations that can effectively evaluate the complexity of their candidate systems early in their life cycle will stand a greater chance of successfully delivering such systems. An effective evaluation can usefully inform important operational and technical decisions, such as what is an appropriate architecture for the proposed system, who are the key stakeholders, what are the key risks and how can they be mitigated, should we even proceed with the project? However, since system complexity can be understood and approached in several different ways, such evaluation is challenging.[9,10]

Investment continues for research into complex systems engineering, for example, £2M of the UK Engineering and Physical Sciences Research Councils's current £7M funding into complexity science research is awarded to the Thales Bristol Hybrid Partnership in Autonomous Systems (Grant EP/R004757/1) to address the challenge of hyrbid autonomous systems engineering. Growth in autonomous systems, smart cities and systems-of-systems deployments foreground the challenges of increasing system complexity, where a large number of diverse and interdependent components, subsystems, and systems interact via nonlinear relationships resulting in emergent behavior and properties that can be difficult to predict and understand, such as the resilience of these complex systems.[11–13]

While the *complexity* of a system is an important characteristic for organizations trying to realize systems, and the term "system complexity" is used frequently, the reality is that this is a contested term, subsuming a myriad of constituent definitions, perspectives, and emphases. The motivation for the study presented here is to explore the extent of the apparent tension between these multiple perspectives on what the term "system complexity" means to the systems engineering community. To do so, we collect judgments of the overall importance of a number of system complexity factors, and also pair-wise comparisons between them, from members of the systems engineering community.

In this study, we define the term "system complexity factors" to refer to a range of different aspects of system complexity that may contribute differentially to systems engineering outcomes (e.g., structural complexity, functional complexity, development complexity). The present study explored systems engineers' views on the absolute and relative importance of these different contributing factors to system complexity for early–life cycle system evaluation. The study data were collected using an online questionnaire of 55 members of the International Council on Systems Engineers (INCOSE) conducted over a four-month period (March to June 2019). Participants were asked to rate the importance of six candidate factors contributing to system complexity (system complexity factors) on a Likert-scale and also via a series of pairwise comparisons. Participants were also asked to rate their prior experience evaluating the same system complexity factors. The participants' experiences of conducting complexity evaluation were also measured and the influence of this experience on system complexity factor importance judgments was assessed.

If the community is using a set of terms relating to "system complexity" in a mature and coherent manner, we would expect the following features to occur in response to our survey. While it might be that some system complexity factors are more important than others, and that some are of roughly equal importance, we would expect to see, at the individual and sample population levels, evidence of coherent mental models of system complexity factor importance. That is, we would expect most respondents to be transitive in their judgments, and for consistency between relative and absolute judgments of importance. Although there may be some inconsistencies in responses, we would expect these to occur for terms that are judged to be of similar importance or of low overall importance. We might also expect to find more experienced practitioners to have more consistent judgments than less experienced practitioners, or for practitioners with similar backgrounds and experiences to share similar judgments.

The purpose of this paper is to collate judgments on the relative and absolute importance of terms relating to "system complexity" in order to explore the maturity of the community's lexicon. To enable this, we collect judgments on system complexity factor importance in an absolute sense by asking for judgments on an ordinal scale and in a relative sense by asking for judgments on pair-wise comparisons. The results are presented as trends in the data when system complexity factors are evaluated on an ordinal scale, trends in the data when system complexity factors are evaluated in a pair-wise comparison, and reflections on free-text answers.

This paper is structured as follows; first, a literature review contextualizes the identified system complexity factors used in the qualitative questionnaire, then the design of the questionnaire is described before discussing the trends in the data. Finally, potential rationales for the results are offered and implications for organizations hoping to better evaluate system complexity are discussed.

2 | LITERATURE REVIEW

A significant challenge for those wishing to evaluate system complexity, and one that persists despite considerable research effort, is finding a single, agreed definition of the term "system complexity" itself.[3,6,14,15] Even determining a distinction between a complex system and a complicated system is not unanimously agreed by the community, where some argue the distinction is between how ordered a system is and therefore how predictable a system is due to the presence (or absence) of nonlinearities and changes within the system, which give rise to emergent behavior[16–18], others emphasize the distinction in terms of how difficult a system is to understand or successfully realize, stressing that complexity is observer dependent.[19]

Some researchers argue that engineering efforts should be concerned, primarily, with *dynamic complexity*[20–22], while others emphasize *sociopolitical complexity*[22,23] or *structural complexity*[3–5,22,24–28] (see also *descriptive complexity*[29]), and others have provided extensive reviews of different definitions further highlighting the diverse conceptual landscape.[30,31] These ideas, and that of Sillitto[22] have been collated into a "Complexity Primer for Systems Engineers"[4] and are also found in the INCOSE Systems Engineering Body of Knowledge for complex systems.[32]

From this myriad of definitions, it is clear that what counts as system complexity is dependent on perspective; on which aspects of a system are deemed important and for what reasons. Further, there are several different *types* of system complexity identified by literature; Fischi, Nilchiani, and Wade draw a distinction between complexity from the perspectives of "the system being observed," "the capabilities of the observer," and "the behavior the observer is attempting to predict"[21] while Simpson and Simpson categorize the complexity of an engineered system as one of the four following types; "cognitive complexity," "behavioral complexity," "organic complexity," and "computational complexity."[33,34]

Moreover, what counts as a reasonable approach to defining system complexity depends on what type of system of interest (SoI) is being considered; is it limited to the *technical system(s) being developed and deployed*, or does it also include the *systems of processes and resources that are involved in developing and deploying such technical systems*?[35] Is the *project* that strives to realize the system under consideration?[36–39] Does it include the processes of utilizing the system once deployed or the user's perceptions of how complex the system is (e.g., how familiar users of the system are with important features of the system)?[40,41] What is the boundary of the SoI; is it the physical context of the implemented system or does it also include the more extended strategic/business context?[42–46] While several approaches purport to provide a quantitative measure of the complexity of a system, they more realistically provide a quantitative measure of the complexity of a particular representation of a system (i.e., a particular view on the architecture of a system).[25,26,47–50] A distinction is also required between the complexity of a representation of a system (i.e., the structural complexity of a system architecture) and the qualitative *perceived* and observer-dependent complexity of the system.[25,26,40] As a consequence, the development of unambiguous and reliable *measures* of system complexity is a considerable challenge.

While several criteria linked to system complexity, such as "requirement difficulty," "cognitive fog," and "stable stakeholder relationships," have been found to have statistically significant correlations with system realization project outcomes, these criteria are far from exhaustive and there is considerable difficulty in accurately measuring them, despite estimating tools.[2,51,52] Further, while metrics exist for quantifying the complexity of software (e.g., cyclomatic complexity[53], lines of code[54]), and conceptually similar metrics exist in the product engineering domain, such as the number and connectivity of physical system components, interfaces, and architecture topology[47,48], when evaluating the complexity of a system architecture[25,26,49,50], developing metrics for a diverse system as a whole remains a challenge.[55]

Despite several decision support or evaluation tools being available to characterize projects and systems[16,43,44,56–58], each of which may be useful for complexity evaluation, there is a challenge in coalescing several perspectives and measures into a coherent whole. One such decision support tool, developed and used by Thales Group, has been previously reviewed and helped to define the identified system complexity factors used here[58–60]. The Thales Group "Complexity Profiler" evaluates the complexity of a candidate system against eight complexity factors which are evaluated on an integer scale (1–4): the impact of the environment on the solution, the stability of the operational concept, user diversity, external stakeholder involvement, life cycle interlacing, systems engineering effort, the stability of system behavior, and the engineering organization.

While the literature surveyed readily acknowledges that there exists a lack of clarity on what the term "system complexity" means to the community, the motivation for this paper is to characterize this lack of clarity, by exploring the extent and nature of the tension between the multiple perspectives advocated for in the literature. Addressing this issue is a step toward addressing the wider question: "To what extent can an organization effectively evaluate system complexity during the early phases of a system lifecycle?"

Based on the literature considered above, we identified six system complexity factors that we use in the qualitative questionnaire. Here, we define each factor and describe their provenance in the literature. First, given the complicated landscape of system complexity factors used by industry and academia, it is necessary to identify and define those used in the survey, rather than exclusively utilize the eight Thales Group factors, as the system complexity factors used by Thales Group do not fully address the aspects identified in the academic literature and often combine multiple terms used by academic literature into a single term. For example, Thales Group use the term "system engineering effort and criticality" to address issues related to the technical novelty of technology and the scale of the development, which in academic literature are two distinct aspects. The six identified factors used here are therefore suggested as a comprehensive amalgamation of identified system complexity factors from both the academic literature and the Thales Group "Complexity Profiler." Their wording is either taken directly from academic literature or modified to ensure the factors consider distinct and unambiguous aspects.

We use the term "Technical Novelty" to represent the number of similar systems the organization has already developed in the same deployment domain, the amount of reuse in the system, the number of high added-value elements, and the level of innovation required to deliver the system. This notion of system complexity has been given multiple different terms: "the difficulty of creation,"[10] the "implementation context and system context,"[43] "socio-political complexity,"[2–4,23,32,61] and "system engineering effort and criticality."[59]

A commonly used term is "Structural Complexity,"[2–4,22,24,32] which we define as the number, diversity, distribution, connectivity, and constraints on constituent components, subsystems, systems, and operational nodes. The term is also related to what has been termed the "implementation context,"[43] and the "system engineering effort and criticality,"[59]

We define "Functional Complexity" as the number, behavior, interdependencies, and synchronization of functions and functional chains, including data types, processing, and memory constraints and algorithms.[58,60] This notion of system complexity can be considered as related to the difficulty conducting functional analysis and allocation[13,62,63].

We use the term "Behavioural Complexity" to mean the ability to define and predict system modes, functions, states, behavior, performance, and missions, including degree of autonomy and the impact of the environment. This is akin to what has been termed "dynamic complexity,"[3,4,20,21,23,32,64] relating to both the "strategic" and "system context,"[43] and has also been termed "operational concept stability" and "system behaviour stability."[58–60]

"Development Complexity" is the amount, and availability, of resources required to develop the system throughout its life cycle, including the interlacing of programs, the degree of challenge of requirements, and the maturity of technology, regulations, standards, processes, and methodologies. Again, this notion of system complexity has been used under various labels: "the difficulty of creation,"[10] "socio-political complexity,"[2–4,23,32,61] It relates to the "system context" and "implementation context,"[43] and has also been termed "development process complexity," "operational complexity," and the "impact of environment on the solution."[59]

Finally, "Organisational Complexity" is the number, diversity, level of support, and involvement of internal and external system stakeholders. This notion of system complexity is akin to the term "socio-political complexity,"[2–4,23,32,61] "life-cycle interlacing," "user diversity," and "engineering organisation"[59] and relates to the "stakeholder context."[43]

3 | METHODOLOGY

The questionnaire contained seven sections: (1) consent to participate; (2) instructions; (3) Likert-scale ratings of the importance of each individual system complexity factor and the opportunity to provide free-text describing any additional important system complexity factors; (4) Likert-scale rating of experience evaluating system complexity factors and a free-text answer of what other factors participants have experience evaluating; (5) pair-wise comparisons between every pair of system complexity factors; (6) free-text describing the participant's experiences of conducting complexity evaluation within their organization; (7) participant background information. The questionnaire instructions, and text before questions, prompted respondents to maintain a single context in their mind for the entirety of the questionnaire (i.e., to consider the questions in the context of a single SoI) to reduce the risk of respondents changing their SoI or context and therefore providing incoherent views.

The order of the test items in Section 3. and Section 4. were fully randomized for each participant. For the pair-wise comparison questions (Section 5.), the presentation order of the pairwise comparisons was shuffled during questionnaire design, and for every respondent the order in which system complexity factors were presented within each individual pair-wise comparison was shuffled.

The questionnaire and resulting data can be found in Ref. 65. Data were collected online between the March 11, 2019, and June 10, 2019. The questionnaire was distributed by email to members of INCOSE UK and published on the news feed of the INCOSE International website.[66] Prompts to complete the questionnaire were provided by social media posts (LinkedIn) on the official INCOSE Group and official INCOSE UK Group, along with emails to members of INCOSE UK and members of the INCOSE Architecture Working Group. Respondents had the option to provide their email address after completing the questionnaire for the chance to win a £50 Amazon gift voucher as an incentive to complete the questionnaire.

After providing electronic informed consent, each participant completed the self-administered online questionnaire. This protocol was approved by the Ethics Committee of the University of Bristol on February 19, 2019 (application ID 81402).

We use Fleiss' κ to measure the degree to which respondents agree on rankings of system complexity factor importance, taking the impact of chance agreements into consideration (see Appendix A).

4 | RESULTS

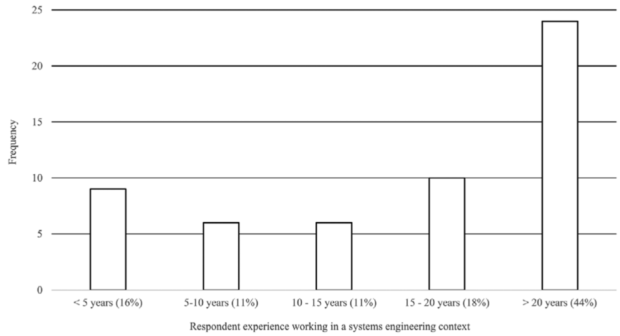
The results of the data analysis on the questionnaire responses are presented in the following manner. First, we describe the make up of the sampled population, before reporting the overall between-participant agreement on the relative importance of system complexity factors. Then, the results are grouped under subheadings in response to specific research questions: How important are different system complexity factors? How are system complexity factors related? Are there distinct views within the participant population? How does experience evaluating a system complexity factor relate to its perceived importance? Are the system complexity factors explored here exhaustive? Are participants internally consistent in their ratings of system complexity factor importance? Are ratings of system complexity factor importance consistent between question types?

The make up of the sampled population in terms of their experience, role, employment sector, and employment location is shown in Figure 1, while Figure 2 shows individual experience conducting system complexity evaluation and the frequency with which their employing organization conducts system complexity evaluation. The sample population includes experienced engineers. Nearly half (44%) of the respondents have over 20 years working in a systems engineering context, only some (16%) were relatively inexperienced. The roles that best describe the respondents were "Systems Engineer" (44% of respondents) and "Systems Architect" (24% of respondents). A range of employment sectors were represented by the sample population, the most frequent sector being "Defence and Space" (38%). The respondents were predominately employed within Europe (76%). When asked how much experience respondents have conducting system complexity evaluation, from options of "Not Sure," "None," "Not A Lot," "Some," "Quite A Lot," and "Lots," the modal responses was "Some" experience. When asked if the organization they are affiliated with conducts system complexity evaluation, from options of "Not Sure," "Never," "Rarely," "Sometimes," "Very Often," and "Always," the modal response was "Sometimes."

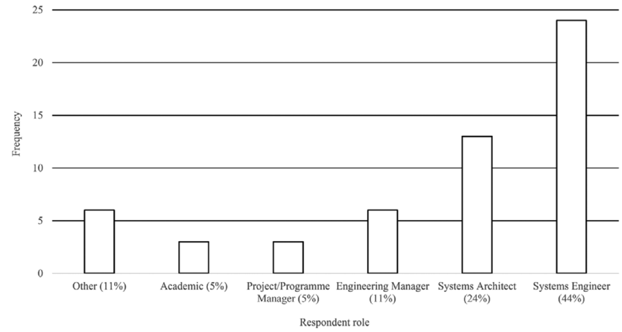
Between-participant agreement on the relative importance of system complexity factors is low; when asked to rate the importance of the six complexity factors on an ordinal (Likert) scale ("Extremely Important," "Moderately Important," "Somewhat Important," "Slightly Important," "Not At All Important"), $\kappa = 0.021$ ($Z = 3.158$, p-value $= 0.002$).

It could be argued that the lack of between-participant agreement on the relative importance of system complexity factors is due to the different backgrounds and experiences participants have had with system development projects. We consider subpopulations based on their responses to self-reported background questions and recalculate Fleiss' κ.

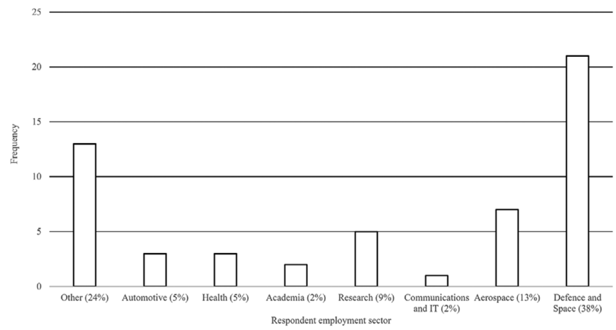
Respondents who reported they had over 20 years of experience working in a systems engineering context had a different ranking of system complexity factor importance, compared with the overall population shown in Table 1, and a lack of between-participant agreement with others of the same experience level ($\kappa = 0.028$, $Z = 1.806$, p-value $= 0.070$). Similarly, for respondents who reported "Systems Architect"



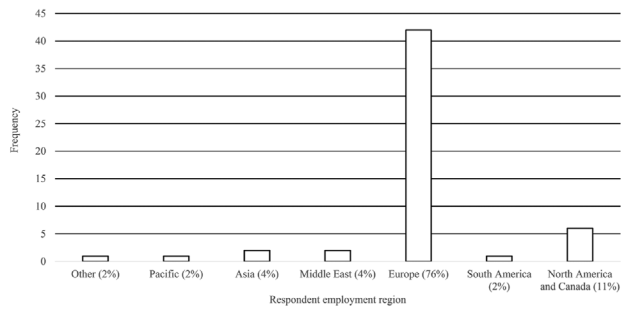
(a) Distribution of answers to the question "Please indicate your experience working in a systems engineering context." No respondents gave the answer "I have not worked in a systems engineering context".



(b) Distribution of answers to the question "Which of the following best describes your current role?" No respondents gave the answer "Student" or "Business Analyst".

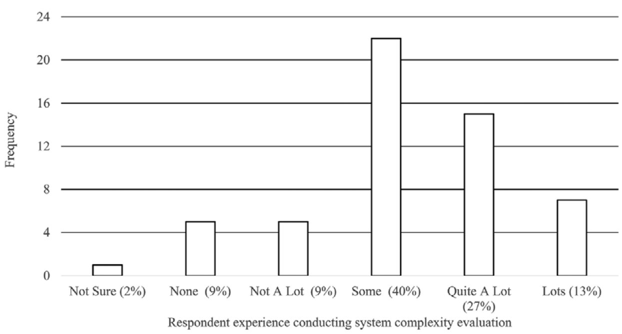


(c) Distribution of answers to the question "Which sector are you currently predominately employed within?".

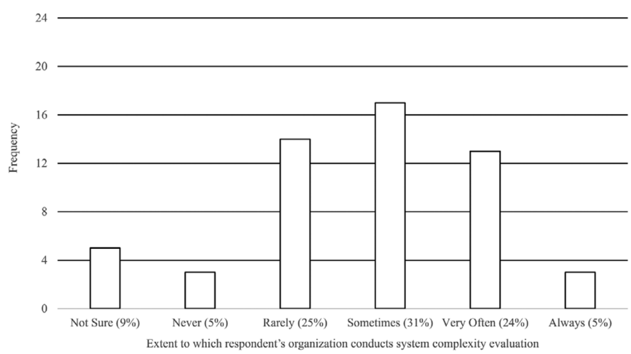


(d) Distribution of answers to the question "Which region do you predominately work in?" No respondents answered "Africa".

FIGURE 1 Distributions of responses for demographics questions (left to right): Experience working in a systems engineering context, current employment role, current employment sector and current employment region



(a) Distribution of answers to the question "How much experience do you have conducting system complexity evaluation?".



(b) Distribution of answers to the question "Does your organization conduct system complexity evaluation?".

FIGURE 2 Distributions of responses to the questions (left to right): "How much experience do you have conducting system complexity evaluations?" and "Does your current organization conduct system complexity evaluation?"

TABLE 1 Mean rank of experience and importance rating for the six complexity factors for the population ($n = 55$), including mean rank of importance rating for those who self-report as highly experienced working in a systems engineering context (> 20 years experience, $n = 24$), and for those who self-report as "Systems Architects" ($n = 13$)

	Population		"Highly Experienced"		"Systems Architect"	
Complexity factor	**Importance**	**Experience**	**Importance**	**Experience**	**Importance**	**Experience**
Organizational complexity	4.01	3.51	4.13	2.98	3.96	2.77
Behavioral complexity	3.91	3.68	4.06	4.00	3.88	3.73
Functional complexity	3.68	3.84	3.42	4.13	3.62	4.08
Development complexity	3.65	3.69	3.88	3.42	3.65	4.15
Structural complexity	3.01	3.37	2.77	3.52	3.73	3.46
Technical novelty	2.75	2.91	2.75	2.96	2.15	2.81

Note: Respondents tend to view technical novelty as the least important aspect when evaluating a novel system to be engineered and organizational complexity as the most important. Experience ratings ranked as 0 = "Not At All Experienced", 1 = "Slightly Experienced", 2 = "Somewhat Experienced", 1 = "Moderately Experienced", 4 = "Extremely Experienced". Complexity factor importance ratings ranked as 0 ="Not At All Important", 1 = "Slightly Important", 2 = "Somewhat Important", 1 = "Moderately Important", 4 = "Extremely Important".

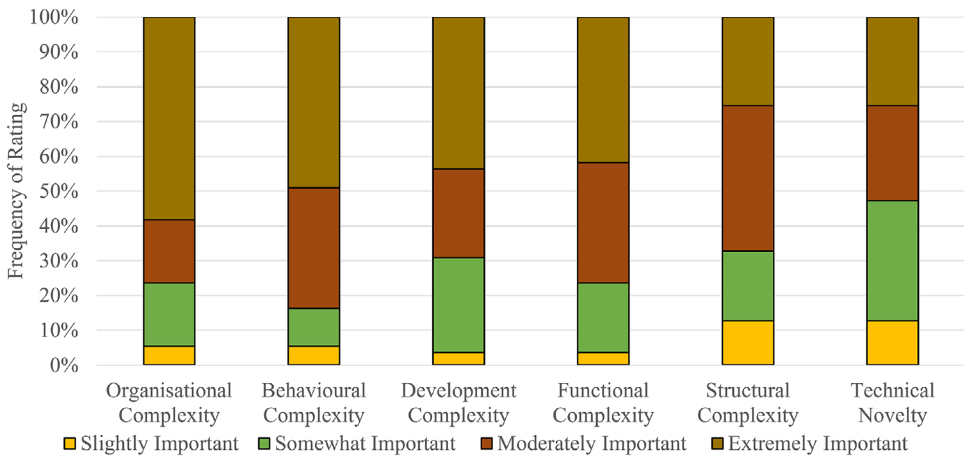


FIGURE 3 Frequency of importance ratings for each complexity factor (n = 55) when respondents were asked to rate the importance of the six complexity factors, shown for each complexity factor. No factors were rated as "Not At All Important"

to be the best role descriptor of their work, $\kappa = 0.008$, $Z = 0.299$ p-value $= 0.764$, indicating a lack of between-participant agreement with others of the same experience level.

4.1 | Importance of different system complexity factors

When respondents were asked to rate the importance of the six system complexity factors on an ordinal (Likert) scale ("Extremely Important," "Moderately Important," "Somewhat Important," "Slightly Important," "Not At All Important"), "Organisational Complexity," "Behavioural Complexity," "Development Complexity," and "Functional Complexity" appear to be particularly important to the community as a whole, shown in Figure 3 and Table 2, with modal ratings of "Extremely Important" for each factor. "Structural Complexity" was not considered to be as important, with a modal rating of "Moderately Important." "Technical Novelty" appeared to be the least important, with modal rating of "Somewhat Important." Respondents who reported "Systems Architect" to be the best role descriptor of their work rated "Structural Complexity" third more important whereas the whole population rated it fifth most important.

We test to see if the responses are essentially random by conducting a χ^2 test on the distribution of responses for each system complexity factor. The results are shown in Table 2 and Table 3. The χ^2 test implies that the null hypothesis that the results are random can be rejected with high confidence, particularly for "Organisational Complexity," "Behavioural Complexity," and "Functional Complexity."

4.2 | Relationships between system complexity factors

Next, we examine correlations (Spearman's rank order correlations, ρ) between the scoring of system complexity factor importance. The results are shown in Table 4 and Figure 4. Note that we are considering 15 correlations here and hence will only be interested in correlation coefficients that are significant at the $p < 0.01$ level or better. Interestingly, "Technical Novelty," with a low median rank and mode, has a significant positive correlation with "Functional Complexity." It makes

TABLE 2 Frequency of importance ratings for each complexity factor (and shown as a percentage of respondents)

Complexity factor	Not at all (%)	Slightly (%)	Somewhat (%)	Moderately (%)	Extremely (%)	χ^2
Organizational complexity	0 (0%)	3 (5%)	10 (18%)	10 18(%)	32 (58%)	34.673***
Behavioral complexity	0 (0%)	3 (5%)	6 (11%)	19 (35%)	27 (49%)	27.545***
Development complexity	0 (0%)	2 (4%)	15 (27%)	14 (25%)	24 (44%)	17.800***
Functional complexity	0 (0%)	2 (4%)	11 (20%)	19 (35%)	23 (42%)	18.818***
Structural complexity	0 (0%)	7 (13%)	11 (20%)	23 (42%)	14 (25%)	10.091*
Technical novelty	0 (0%)	7 (13%)	19 (35%)	15 (27%)	14 (25%)	5.436

Note: Corresponding χ^2 test statistic (d.f. = 3) and p-value against an equal distribution of ratings for each Complexity Factor with "Not At All Important" removed because no respondent rated any of the complexity factors as "Not At All Important," where *** denotes $p < 0.001$, ** denotes $p < 0.01$, and * denotes $p < 0.05$, otherwise not significant. When tested against an equal distribution of ratings with "Not At All Important" included, all χ^2 p-values are significant ($p < 0.001$).

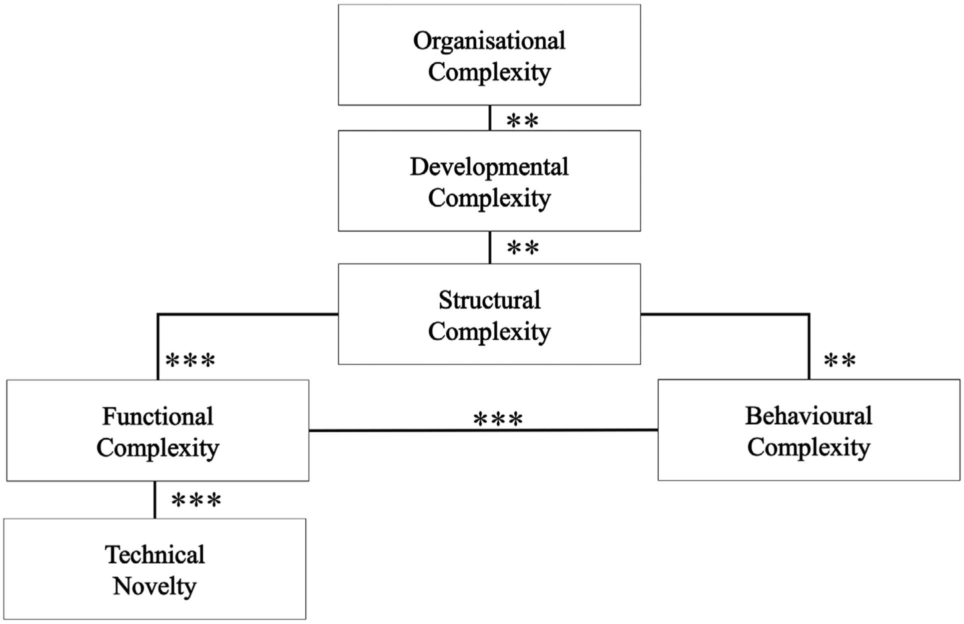


FIGURE 4 Spearman's rank order correlation coefficients ($\rho(55)$) between complexity factors for the sample population, * corresponds to $p < 0.05$, ** corresponds to $p < 0.01$ and *** corresponds to $p < 0.001$. All correlations shown are positive

TABLE 3 Frequency of ratings with corresponding χ^2 test statistic (d.f. = 1) and p-value against an equal distribution of ratings for each Complexity Factor, with "Slightly Important" and "Somewhat Important" collapsed into the term "Average Importance" and "Moderately Important" and "Extremely Importance" collapsed into the term "Particular Importance"; Again, *** denotes $p < 0.001$,** denotes $p < 0.01$, and * denotes $p < 0.05$, otherwise not significant

Complexity factor	"Average" ("Slightly" + "Somewhat")	"Particular" ("Moderately" + "Extremely")	χ^2
Organizational complexity	13	42	15.291***
Behavioral complexity	9	46	24.891***
Development complexity	17	38	8.018**
Functional complexity	13	42	15.291***
Structural complexity	18	37	6.564*
Technical novelty	26	29	0.164

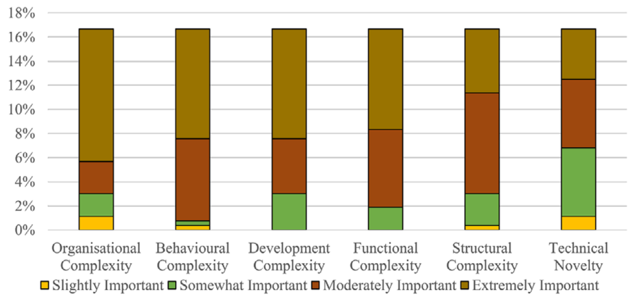
sense that the community relates "Organisational Complexity" and "Developmental Complexity" as both terms relate to the *system which develops the SoI*. Similarly, "Functional Complexity," "Structural Complexity," and "Behavioural Complexity" all relate to the *technical SoI to be developed*. The two subpopulations examined earlier ("experienced participants" and "systems architects") have similar judgments to the overall sample population on which system complexity factors are related together, with have no new factor–factor correlations arising.

4.3 | Distinct views within the participant population

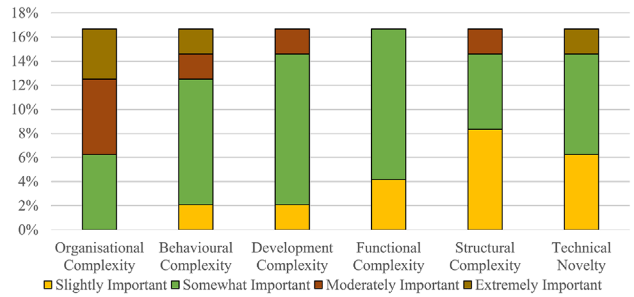
In this section, we examine whether there are distinct clusters of the population that share the same judgments on system complexity factor importance. The data were analyzed using a hierarchical

TABLE 4 Table of Spearman's rank order correlation coefficients ($\rho(55)$) between complexity factors when respondents were asked to rate their importance on a Likert-scale ($d.f. = 53$); $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$, otherwise ρ values are not significant

	Development complexity	Structural complexity	Organizational complexity	Behavioral complexity	Technical Novelty	Functional complexity
Development complexity	-					
Structural complexity	0.334**	-				
Organizational complexity	0.426**	−0.051	-			
Behavioral complexity	0.201	0.407**	0.060	-		
Technical novelty	0.200	0.174	−0.108	0.249	-	
Functional complexity	0.251	0.602***	−0.029	0.461***	0.450***	-



(a) Cluster A's (N = 44, 80% of sample population) distribution of complexity factor importance by importance category.



(b) Cluster B's (N = 8, 15% of sample population) distribution of complexity factor importance by importance category.

FIGURE 5 Responses to Likert rating of the importance of complexity factors given by importance category for two identified clusters: A (left, $N = 44$, 80% of sample population), B (right, $N = 8$, 15% of sample population). Cluster A has κ of 0.044, and for a χ^2 test of distribution of importance ratings across two categories ("average importance" and "particular importance") all complexity factors be considered of "particular importance" apart from "Technical Novelty" (statistically insignificant test result). Cluster B has κ of 0.002, and for a χ^2 test of distribution of importance ratings across two categories ("average importance" or "particular importance") all of the complexity factors could be considered to be of "average importance" ($p < 0.05$) apart from "Organisational Complexity" and "Behavioural Complexity" (statistically insignificant test results)

agglomerative clustering algorithm where importance ratings were converted to integers. A number of metrics could be used to examine "distance" between different respondents' views of system complexity factor importance. Here, we use the simplest approach (Manhattan distance). Other approaches could account for individuals who agree on which factor is most important, but disagree on which factor is least important, or those in the "middle of the pack." However, here we focus on the simplest method. From the cluster analysis, flat clusters were determined which grouped the respondents into one of four groups. The resulting clusters ($A - D$) represent 80%, 15%, 4%, and 2% of the sample, respectively. The distribution of ratings within these clusters is shown in Figure 5 and in Table 5.

For Cluster A (80% of sample population, $n = 44$), Fleiss' κ = 0.043, Z = 4.691, p < 0.001, again showing a lack of between-participant agreement on the relative importance of system complexity factors. A χ^2 test was conducted on the distribution of importance ratings across two categories (*"average importance"* or *"particular importance"*), the results of which show that all but "Technical Novelty" could be considered of "particular importance" to this cluster. This large subpopulation appears to view all of the system complexity factors, apart from "Technical Novelty," as being important when evaluating system complexity, but do not agree on the relative importance of the factors, with modal rating of "Extremely Important" for the remaining factors apart from "Structural Complexity," which has a modal rating of "Moderately Important." This subpopulation has a similar judgment to the overall sample population on which system complexity factors are related, with no new factor–factor correlations arising.

For Cluster B (15% of sample population, $n = 8$), Fleiss' $\kappa = 0.019$, $Z = 0.370$, $p = 0.711$, again showing a lack of between-participant agreement on the relative importance of system complexity factors. For a χ^2 test of distribution of importance ratings across two categories (*"average importance"* or *"particular importance"*), all of the system complexity factors can be considered to be of "average importance" ($p < 0.05$).

Based on this analysis, it seems there are distinct clusters within the sample population with two competing views: a large proportion of the

TABLE 5 Mean rank of experience and importance rating for the six complexity factors for the population (n = 55), cluster A (n = 44) and Cluster B (n = 8)

	Population		Cluster A		Cluster B	
Complexity factor	**Importance**	**Experience**	**Importance**	**Experience**	**Importance**	**Experience**
Organizational complexity	3.51	4.01	3.61	3.86	3.31	4.94
Behavioral complexity	3.68	3.91	3.59	3.93	3.88	3.69
Functional complexity	3.84	3.68	3.76	3.76	4.19	2.94
Development complexity	3.69	3.65	3.75	3.72	3.94	3.63
Structural complexity	3.37	3.01	3.39	3.13	3.31	2.81
Technical novelty	2.91	2.75	2.90	2.60	2.38	3.00

Note: Cluster A and B have a similar ranking with the overall population for experience evaluating each complexity factor. Cluster A has the ranking of the top two important complexity factors reversed, with "Behavioural Complexity" rated as the most important and "Organisational Complexity" as the second most important. Conversely, Cluster B has different importance rankings for all the complexity factors apart from the top two most important. Experience ratings ranked as 0 = "Not At All Experienced", 1 = "Slightly Experienced", 2 = "Somewhat Experienced", 1 = "Moderately Experienced", 4 = "Extremely Experienced". Complexity factor importance ratings ranked as 0 = "Not At All Important", 1 = "Slightly Important", 2 = "Somewhat Important", 1 = "Moderately Important", 4 = "Extremely Important".

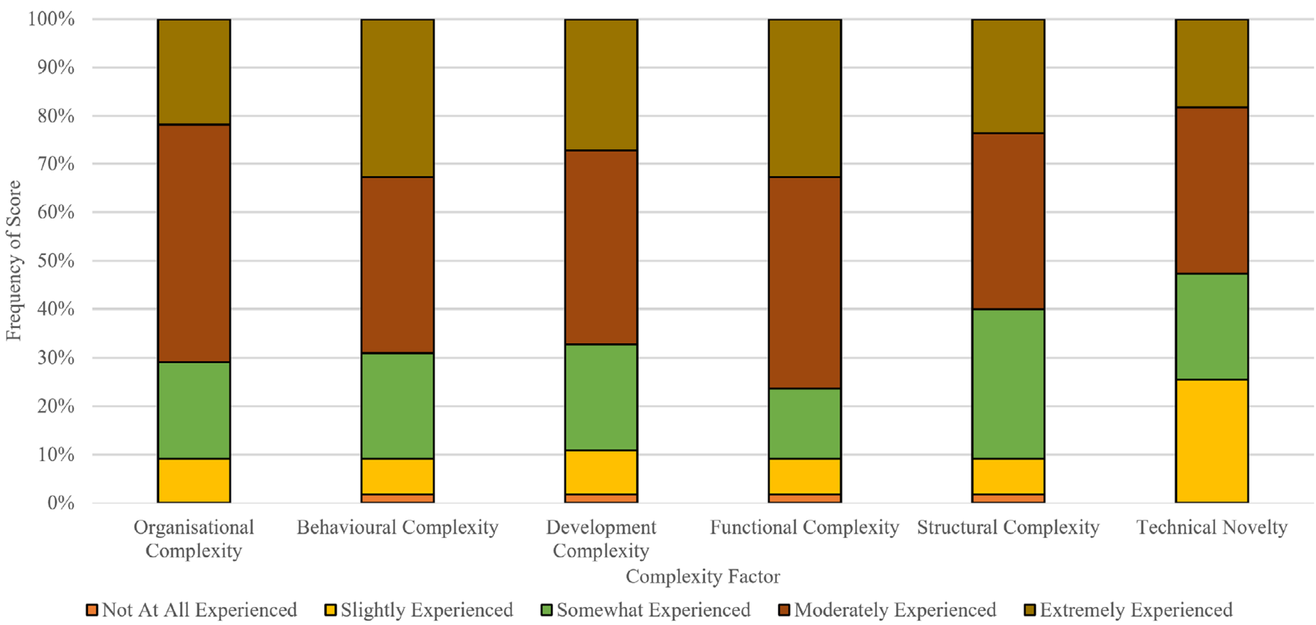


FIGURE 6 Responses to Likert-type rating of respondent's experience evaluating complexity factors with responses shown for each complexity factor

sample population who do not agree on the *relative* importance of these factors but agree on the *absolute* importance of all factors apart from "Technical Novelty," contrasted with a smaller cluster who suggest a lack of *absolute* importance of all of the factors but also do not agree on the *relative* importance of the factors.

4.4 | Relationship between experience evaluating a system complexity factor and its perceived importance

We asked "To what extent do you have experience evaluating the following aspects?" with options of "Not At All Experienced," "Slightly Experienced," "Somewhat Experienced," "Moderately Experienced," or "Extremely Experienced" for each of the six complexity factors. The results are shown in Figure 6.

We collate the ratings of system complexity factor importance and examine the Spearman's rank order correlation coefficient (ρ) with collated ratings of experience evaluating that factor. Overall, $\rho(330) = 0.334$ ($d.f. = 328, p < 0.001, \alpha = 0.01$), demonstrating that generally respondents rate system complexity factors that they have experience evaluating as more important than those that they have less experience evaluating, shown in Table 6. We also examine correlations between the ratings of experience evaluating each system complexity factor, finding that generally experience evaluating one system complexity

TABLE 6 Table of Spearman's rank order correlation coefficients

	Development complexity	Structural complexity	Organizational complexity	Behavioral complexity	Technical novelty	Functional complexity
Development complexity	(0.423)**					
Structural complexity	0.669***	(0.201)				
Organizational complexity	0.377**	0.359**	(0.162)			
Behavioral complexity	0.548***	0.633***	0.348**	(0.302*)		
Technical novelty	0.273*	0.256	0.197	0.349*	(0.579***)	
Functional complexity	0.429**	0.595***	0.252	0.685***	0.357**	(0.143)

Note: Leading diagonal shows Spearman's rank order correlation coefficients ($\rho(330)$) between complexity factor importance and experience evaluating that complexity factor; $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$, otherwise ρ values are not significant. When the ratings of complexity factor importance and experience evaluating that factor are collated into two variables, the overall correlation is $\rho = 0.334$ ($d.f. = 328, p < 0.001, \alpha = 0.01$) demonstrating that generally respondents rate complexity factors that they have experience evaluating as more important than those that they have less experience evaluating. Off-diagonals show Spearman's rank order correlation coefficients ($\rho(55)$) between complexity factors when respondents were asked to rate their level of experience evaluating each factor on a Likert-scale ($d.f. = 53$); $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$, otherwise ρ values are not significant.

factor seems correlated to experience evaluating every other, apart from "Technical Novelty."

4.5 | Relevance of the system complexity factors used in the questionnaire

We check the relevance of the six system complexity factors by examining if the community thinks these six factors are all unimportant. None of the respondents gave any of the six system complexity factors an importance rating of "Not At All Important." Although this does not mean that the six system complexity factors chosen in this survey are exhaustive, they are at least relevant.

We asked each respondent "What other aspects are important to you when evaluating system complexity?" and received a mixture of free-text responses, suggesting that there is a wide range of contextually relevant aspects that are important when evaluating system complexity. The most frequent emerging themes include system interfaces and dependencies (nine responses), nonfunctional requirements including safety and security (eight responses), and client/customer/user complexity (e.g., their understanding of the system, novelty of the system to them, willingness to accept change) (seven responses). We also find further evidence supporting the relevance of the six terms used in the survey as seven respondents answered that these factors were sufficient. When we asked "When evaluating system complexity, what other aspects do you have experience with?" a range of answers were received, suggesting a wide range of aspects that are currently evaluated. The usefulness of these aspects was not reported however. The most frequent answers included nonfunctional requirements (including safety, security, and regulatory compliance requirements) and the "ilities" (e.g., flexibility, adaptability) (seven respondents), financial and commercial complexity (six respondents), and stakeholder complexity (diversity, expectations) (three respondents), while seven respondents answered with no other aspects.

We also asked each respondent to "Please describe your experience of complexity evaluation (for example; the extent to which this type of activity has been a part of your job, the purpose of any complexity evaluation that you have been involved in, how successful or otherwise you felt complexity evaluation was, the challenges you faced, etc.)." While the most common response (23 respondents) was to provide no answer, the second most frequent answer (seven respondents) related system complexity evaluation to risk evaluation (technical, project/program). Answers relating complexity evaluation to risk suggested complexity evaluation are "performed to understand the program risks," "to identify where we carry our biggest risks," "highlights the complexity and its associated risks to leadership," while another respondent answered "what seems to me to be important is to understand what the complexities are - so identification rather than evaluation - and then what the risks /potential impacts associated with those complexities are - and then take management action…" Four respondents answered that a subjective evaluation has been done, where "nothing formal" was done, with a respondent answering that "Evaluation has been on a 'gut' basis; I haven't used any systematic approach," and another respondent answered that complexity evaluation "is often completed at a high level view and too often based on the experience of

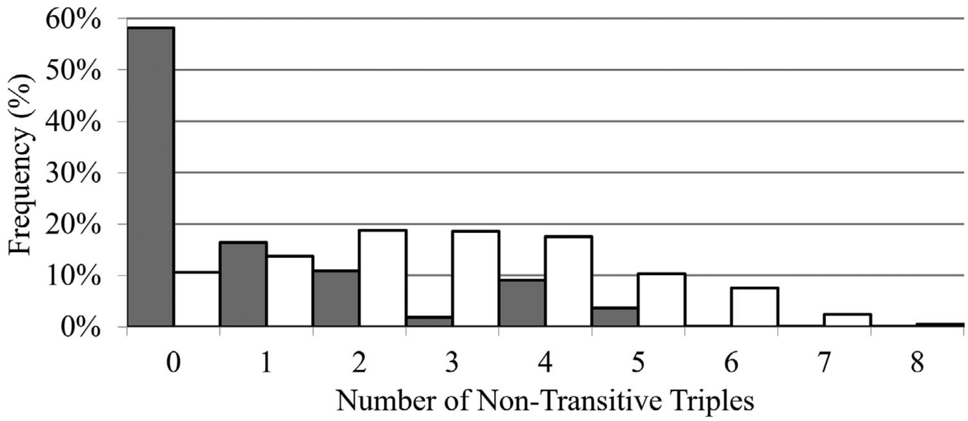


FIGURE 7 Distribution of the number of non-transitive triples ($N = 55$) for the survey respondents (filled bars), where 58% of respondents gave at least one non-transitive response, compared with the distribution of the number of non-transitive triples from a null model ($n =$ 10,000,000) (open bars). A two sample Kolmogorov–Smirnov test was used to test whether the null model and the survey results came from the same distribution. Here, $D = 0.502$, $p < 0.001$ at $\alpha = 0.05$ giving confidence the two samples are not the same distribution

the assessor making it a subjective process rather than objective. Complexity is in the eye of the beholder." These answers provide evidence that some systems engineering practitioners are already using the term "system complexity" as a proxy for system risk (whether technical or programme) but that some lack formal approaches.

4.6 | Participant internal consistency in ratings of system complexity factor importance

An alternative to rating each system complexity factor's importance on a Likert-scale is to elicit pair-wise comparisons between system complexity factors. Pair-wise comparisons may be nontransitive, where a respondent rates system complexity factor A as equally or more important than B, and rates B to be equally or more important than C, but rates C to be equally important or more important than A. By counting the number of nontransitive triples in the pair-wise comparisons, we have an approach to characterize how inconsistent respondents were in their answers. We use the procedure from Ref. 67 to count the nontransitive triples.

Figure 7 shows that over half of respondents (32, 58%) were fully transitive in their responses, 16% gave one nontransitive triple, and 11% gave two nontransitive triples. As there are six system complexity factors, there are a total of 20 triples in a network representing system complexity factors. A null model can be used to determine how many nontransitive triples would be expected to arise if each possible pair-wise comparison between six elements were generated at random. We use a similar procedure to Ref. 67 to create our own null model of non-transitivity expected at random, including the possibility that two factors may be rated as equally important. For this null model, we simulate 10,000,000 sets of responses where each pair-wise comparison between A and B has a 0.36 chance of favoring A, a 0.36 chance of favoring B, or a 0.28 chance of rating them equal, since 228 of the 825 pair-wise comparisons made by participants (28%) were rated as equal. The largest number of nontransitive triples for any respondent was five, compared with eight nontransitive triples that could be expected from the null model. Overall, the population provided fewer nontransitive responses than would be expected by chance, supporting the argument that the majority of the community form their own coherent mental models of system complexity factor relative importance. A two sample Kolmogorov–Smirnov test was used to test whether the null model and the survey results came from the same distribution; $D = 0.502$, $p < 0.001$ at $\alpha = 0.05$, giving confidence the two samples are not the same distribution.

We compare the results of a single randomly selected simulation run (55 sets of responses) with the survey responses, Figure 8, which shows that while there are some triples that are more likely to be answered in a nontransitive way by the respondents, this distribution could be explained by chance selection.

Other than a random distribution of nontransitive triples, we could imagine the most frequent nontransitive triples to be those that are rated to be of similar importance. Surprisingly, the most frequent nontransitive triples include both system complexity factors that are on average particularly important but also those that are not considered to be as important (e.g.,"Structural Complexity" contrasted with "Technical Novelty"). We count the number of nontransitive responses that occur for each possible pair of system complexity factors and examine the correlation between this count, and the difference in mean rating for the same pair of system complexity factors, finding no significant correlation suggesting the nontransitive responses are not systematic; that they are not entirely explained by similarity between judgments of system complexity factors.

4.7 | Consistency between ratings of system complexity factor importance between question types

Judgments of system complexity factor importance should not change depending on whether the sample population were asked to rate their

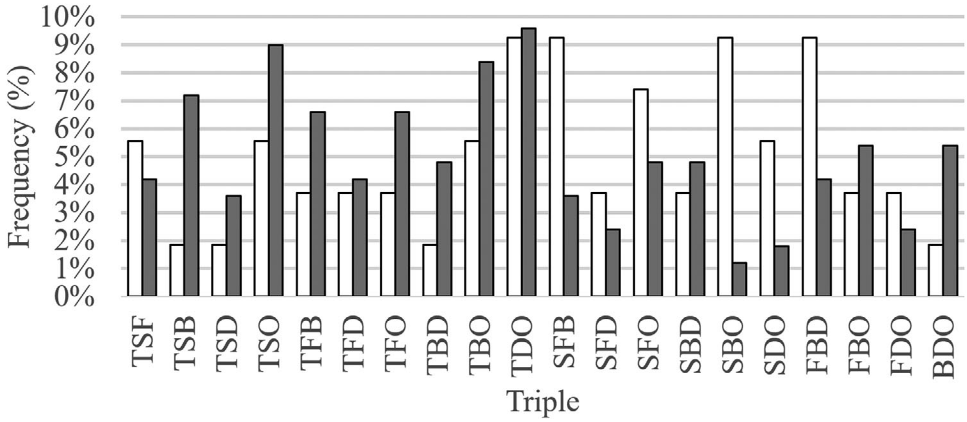


FIGURE 8 Distribution of non-transitive responses across all possible triples for the sample population ($n = 55$, filled bars) and for one run of the null model simulation ($n = 55$, open bars). Labels correspond to: T - Technical Novelty, S - Structural Complexity, F - Functional Complexity, B - Behavioral Complexity, D - Development Complexity, O - Organizational Complexity

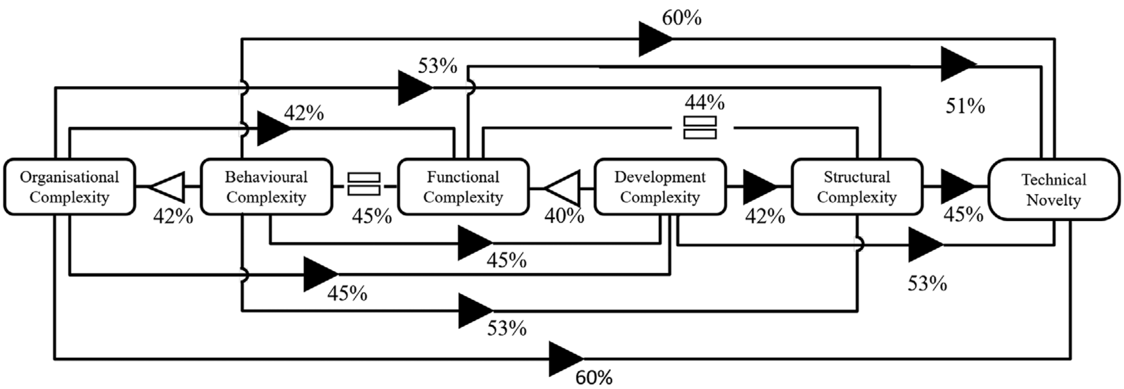


FIGURE 9 The proportion of participants that rated each factor as more or equally important than each of the others. The majority proportion is reported for each relationship. An arrow from complexity factor A to complexity factor B means that more participants judged A to be more important than B than vice versa, with the percentage value showing what proportion of participants judged A to be more important than B. Solid arrowheads are consistent with the overall population judgments. Open arrowheads represent judgments that are intransitive with respect to the overall population judgments. An equals symbol represents majority judgments of equal importance

importance on a Likert-scale or in a pair-wise manner. The proportion of participants that rated each system complexity factor as more important, or equally important, than each of the others is shown in Figure 9 where the highest frequency of responses is shown for each relationship. Interestingly, the population is generally consistent over longer "distances" between system complexity factors in terms of their importance, and inconsistencies are generally found for the highly important system complexity factors. Neither Cluster A nor Cluster B from Section 4.3 produced significantly different results when compared with the overall sample population.

Similarly, Figure 10 shows the discrepancy between the two question types, where the order of system complexity factor importance is different depending on whether the sample population is providing judgments on a Likert-scale or in a pair-wise comparison. The sample population had agreement between the two question types in rating "Organisational Complexity" and "Behavioural Complexity" as being particular important. We compared the responses between the two question types by aggregating the Likert-scale responses (by taking the normalized mean rank) and aggregating the pair-wise comparison responses (by using the procedure from the Analytic Hierarchy Process [AHP] to calculate the aggregate of individual judgment[68]), normalizing the results between 0 and 1. For further details, see Appendix B.

5 | DISCUSSION

Before discussing the implications of the results and limitations of the research, we briefly summarize the main results found during the analysis of survey responses. All six of the terms used here relating to "system complexity" are relevant, but they are not an exhaustive list. The sample population participants identify the absolute and significant importance of "Organisational Complexity," "Behavioural Complexity," and "Structural Complexity" but do not exhibit significant agreement as to the relative importance of these factors. We found two

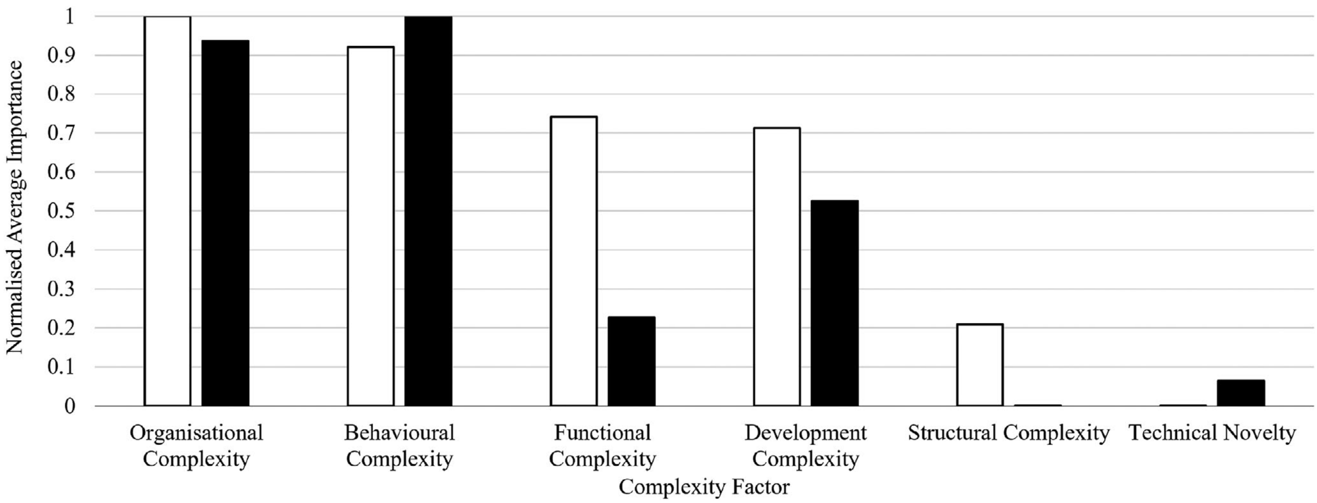


FIGURE 10 Comparisons of responses to the ordinal scale ratings of complexity factor importance (normalized average of ranks, open bars) with responses to the pair-wise comparisons (normalized average of aggregate of individual judgment, closed bars)

competing views within the sample population, a large majority who judge that all six system complexity factors are absolutely important, but did not agree on their relative importance, and a small minority who judge that all six system complexity factors are of average importance, again not agreeing on the relative importance of the factors. Considering more homogeneous subpopulations did not increase the amount of agreement on the relative importance of system complexity factors. Several system complexity factors are considered to be related to each other by the sample population, for example, "Organisational Complexity" and "Developmental Complexity" are both terms which relate to the *system which develops the SoI*, and "Functional Complexity," "Structural Complexity," and "Behavioural Complexity" are all terms that relate to the *technical SoI to be developed*. These correlations are stable across subpopulations. Generally respondents rate system complexity factors that they have experience evaluating as more important than those that they have less experience evaluating, although this is not true for all individual system complexity factors. Most of the sample population gave fully transitive responses when asked to evaluate the importance of the system complexity factors in a pair-wise manner, with low nontransitive of responses. Where there were nontransitive responses, it was not systematic and it is not possible to rule out that there was noise in the results.

Inconsistencies appear for the sample population when comparing the results of pair-wise comparisons with judgments on an ordinal scale, suggesting that while the majority of respondents have coherent and consistent mental models of system complexity factor importance, the community overall could stand to improve their mental models of system complexity factor importance. It likely remains that the perceived relevance and importance of system complexity factors is strongly linked to individual experiences on system development projects, it remains that the community could improve the consistency of their judgments on the relative importance of these factors by conducting more frequent, formal evaluations.

The nontransitivity found may be due to inconsistencies in respondent's mental model of system complexity factor importance. Perhaps because the SoI they maintained in their mental model changed over the course of completing the questionnaire, or similarly, perhaps the context they imaged themselves located within in their mental model changed over the course of completing questionnaire, or finally perhaps they choose different aspects of the system complexity factor definition to "focus on" while completing the questionnaire. While the questionnaire was designed to prompt the respondents to maintain the same SoI and context over the course of the questionnaire it remains that this issue may contribute to the inconsistency in results. However, neither the distributions of nontransitive responses, nor the distributions of importance ratings, were randomly distributed or particularly noisy, instead individuals hold different views to one another.

It may simply be that the sample population lacks stronger consensus on the importance on system complexity factors precisely because the population is diverse; with widely different mental models of their SoIs, operating in different business and operational contexts, within different domains and at different levels of abstraction. Support for this argument is in the transitive responses and the consistent correlations between factors and the correlation between experience and ratings. This position is consistent with the fact that the literature has a diverse set of definitions of system complexity. Even if a definition of system complexity is provided to the community, the *interpretation* of this definition in diverse contexts may create the range of responses found here.

The results of the questionnaire support the literature surveyed in that the term "system complexity" is ambiguous and contextually sensitive to systems engineers, where diversity in respondent roles, experience, operating domain, expertise, their systems of interest, the context their SoI operates within, etc., hinder an unambiguous and consensus view on the importance of factors that contribute to system complexity.

Several implications for organizations wishing to evaluate the complexity of their systems derive from the findings that (i) the six terms used here are relevant, but not exhaustive, and (ii) that individuals have coherent and consistent mental models but that the community as a whole does not. First, in the absence of a decision support tool that explicitly examines system complexity, organizations have an opportunity to create such a tool that asks stakeholders to explicitly evaluate the complexity of their systems, using the six system complexity factors used here as an initial prompt. Similarly, organizations that currently utilize such a process or tool should ensure that their tooling includes consideration of the six factors considered here. During evaluations, particular attention should be given to evaluating the organizational complexity, behavioral complexity, and structural complexity of candidate systems. The perceived importance of structural complexity is reflected by a wealth of academic literature on evaluating the structural complexity of a system or product architecture. Similarly, the literature foregrounds the challenges and implications of autonomy on complex systems engineering where autonomy can be considered as an archetype of behavioral complexity.[11,69,70] Further, recent research[71] has suggested that despite being a widely used term, the technology readiness level (TRL) of a target system presents significant challenges in evaluating the difficulty of the associated systems engineering project, which is supported here given the lower importance given to technical novelty.

There are several cautions and considerations that an organization or systems engineer evaluating the complexity of their systems needs to be cognizant of. First, they should note that the six factors used here are not exhaustive, and additional factors are likely to emerge and to evolve over time. Second, complexity evaluation requires an acknowledgment that the importance of system complexity factors appears to be dependent on observer perspective. While a systems engineer or systems architect might consider a particular line of inquiry to be a highly important activity worth investing resources into (e.g., explicitly determining how structurally complex their candidate architecture is), they must be aware that others may not have the same view, further mired by the contested, ambiguous definitions of the terms often used relating to system complexity.

The results presented here are limited by the fact that we considered six specific complex systems factors and pooled results from a set of systems engineers working across distinct types of engineering sector and project. Further work could consider a wider range of complex systems factors and explore them in the context of a more explicit set of different systems engineering contexts. Over time, organizations may develop their own ontology of relevant system complexity factors: which factors relate to which factors, a potential hierarchy of terms, a distillation of broader terms relating to system complexity into more quantifiable or atomic terms, and how system complexity factors relate to system type, a consideration that was not explicitly examined in this study. Further, they could consider a through-life cycle perspective, whereas this study emphasized the early system life cycle implications of system complexity. Future work should investigate these points further and develop a richer ontology of system complexity factors, one that moves beyond *perceived* importance of system complexity factors and instead seeks unambiguous objective measures that differentially impact on system development projects along with a framework to support the through-life cycle evaluation of system complexity, sensitive to the impact of system type on system complexity.

There are inherent limitations to the questionnaires as a research instrument: first, the Likert-type scale assumes linearity of responses, which may not strictly be true. Second, there may be a fatigue effect while respondents completed the questionnaire, which may have contributed to the nontransitivity found. Although, as many respondents were consistent in their answers and overall the importance ratings of system complexity factors do not appear to be rated at random (tested using a χ^2 test), there can be some confidence in the results despite any fatigue effects. The questionnaire also used pair-wise comparisons to mitigate this concern, which in theory offer a lower cognitive burden for respondents, although the number of individual comparisons required of each respondent may contribute to the fatigue effect.

6 | CONCLUSION

This research has sought to address the question: “To what extent can an organization effectively evaluate system complexity during the early phases of a system lifecycle?” Here, we examine the judgments of systems engineers on the importance of six different factors, which may contribute to system complexity, revealing a lack of significant consensus on which aspects of system complexity are most important when engineering a novel system.

The between-participant agreement on the relative importance of system complexity factors is low, $\kappa = 0.021$, $Z = 3.158$, p-value $= 0.002$. In terms of absolute importance, the overall participant population rated “Organisational Complexity,” “Behavioural Complexity,” and “Functional Complexity” as particularly important but did not rate “Development Complexity,” “Structural Complexity,” and “Technical Novelty” as particularly important. However, the overall participant population includes two competing views: a majority view that all of the factors are important, but with no agreement on relative importance amongst them, contrasted with a minority view that the terms are only of average importance but again with no agreement on the relative importance among them. Self-reported demographics do not appear to explain the variation in views. Several system complexity factors are considered to be related to each other, for example, “Organisational Complexity” and “Developmental Complexity” are seemingly related terms, and “Functional Complexity,” “Structural Complexity,” and “Behavioural Complexity” are seemingly related terms, with the same correlation structures stable across subpopulations. Generally respondents rate system complexity factors that they have experience evaluating as more important than those that they have less experience evaluating, although this is not true for all individual system complexity factors.

The majority of respondents gave fully transitive responses when asked to evaluate system complexity factor importance in a pair-wise manner, indicating a maturity in respondent's mental models of the construct of the term *system complexity* and the overall level of nontransitivity in participant judgments was low (16% of responses were nontransitive). Where nontransitive responses were given, they were not fully explained by factors involved being judged to have similar level of importance. Inconsistencies were found when comparing the sample population judgments of system complexity factor importance on an ordinal scale with judgments provided in a pair-wise manner, suggesting that while individuals have mature mental models, the community could improve the consistency of their judgments. While it likely remains that the importance of system complexity factors depends strongly on individual experience, with the community suggesting a rich set of such relevant factors, consistency in judgments of system complexity factor importance can be improved through more frequent, formal evaluations.

This paper has begun to characterize the community's understanding of its own complex systems vocabulary. Its results suggest that while individual practitioners each tend to hold a coherent view of complex systems factors, aggregating these views does not result in a single coherent consensus.

The fact that systems engineers are not aligned with each other on how important system complexity factors are, and may not appreciate the extent or nature of these misalignments, may hinder efforts to effectively identify, evaluate, and manage system complexity during the early phases of a system life cycle and beyond.

By contrast, if the systems engineering community did agree unanimously on the relative and absolute importance of a set of complex systems factors, effective strategies could be developed to include system complexity as a system architecture evaluation criterion, allowing one architecture to be assessed as more desirable than another due to its lower overall complexity.

Alternatively, even if the community continues to maintain multiple different positions on the relevance and important of different complex systems aspects, if these different positions are held by different individual practitioners for good (context-specific, empirically supported) reasons and these differences are well recognized, well understood, and well articulated, this would also enable complex systems evaluation to be undertaken profitably (although the process would be more onerous) and allow effective trade-offs to be made during a project's design phase or later.

However, given the lack of a consensus view on the relative and absolute importance of the terms explored here, systems engineers, architects, and organizations currently are left without clear guidance on which features of system complexity to pay particular attention to. While the results of this study cannot be used to provide a full model of complexity factor importance that is universally applicable to all in the community, the results of the survey can instead be used to make recommendations on mitigating the challenges presented by the revealed ambiguity of system complexity terms.

First, care should currently be taken when using terms related to *subcomponents* of system complexity because they remain open to interpretation and do not automatically avoid the ambiguity that is recognized to be associated with the overarching term "system complexity" itself. Organizations wishing to evaluate system complexity should work with a set of clear definitions of the terms that they use, defined in such a way as to address relevant subcomponents of complexity.

This task of defining complexity-related terms should take into consideration the type of system under evaluation and the contextual factors that are most relevant to the evaluation. For instance, the complexity of a predominately mechanical system's architecture may require a different language from that appropriate to the evaluation of a predominately software-based system's architecture, even if the terms being used appear to be the same.

Moreover, combining evaluations of subsystems to achieve an evaluation of overall system complexity should not be regarded as a simple process of addition. For example, the super-system composed by combining the mechanical system with its relevant software-based system may raise entirely new issues that require careful evaluation and may trigger re-evaluation of the original subsystems. It is at this point that different interpretations or assessments of key terms and concepts can cause most damage.

Consideration should also be given to the complexity of a candidate system throughout its life cycle, rather than relying solely on early-phase evaluations. Care should be taken to consider that the relative importance of different system complexity factors may change as the system progresses through its life cycle. Finally, consideration should be given to ensure that a closed set of system complexity factors is used during evaluations.

Foregrounding the current lack of consensus on the factors implicated in system complexity also provides an opportunity for the community to direct future research toward the development of a holistic framework to support the evaluation of system complexity.

ACKNOWLEDGMENTS

The authors would like to thank Dave Harvey and Jean-Luc Garnier at Thales for their insights, guidance, and support. The authors also wish to thank the anonymous reviewers for their insightful feedback and contributions.

ORCID

Matthew W. Potts https://orcid.org/0000-0002-4266-0862

REFERENCES

1. Hartmann R, Belhoff B, Oster C, et al. *A World in Motion, Systems Engineering Vision 2025* [Journal Article]. San Diego, CA: INCOSE; 2014.
2. Sheard SA. *Assessing the impact of complexity attributes on system development project outcomes*. Stevens Institute of Technology; 2012.
3. Sheard SA, Mostashari A. 7.3. 1 A complexity typology for systems engineering. In: *INCOSE International Symposium*. vol. 20. Wiley Online Library; 2010. pp. 933–945.
4. Sheard S, Cook S, Honour E, et al. A complexity primer for systems engineers [Journal Article]. white paper, INCOSE Complex Systems Working Group. 2015.

5. Sheard SA. 5.2.1 Systems Engineering Complexity in Context. In: *INCOSE International Symposium*. vol. 23. Wiley Online Library; 2013. pp. 1145–1158.
6. Potts M, Sartor P, Johnson A, Bullock S. Hidden structures: using graph theory to explore complex system of systems architectures. In: Chapoutot A, Krob D, Roussel A, Stephan F, eds. *Complex Systems Design & Management*. Paris, France: CESAM Community; 2017:117–131.
7. Potts MW, Sartor P, Johnson A, Bullock S. A network perspective on assessing system architectures: foundations and challenges. *Syst Eng*. 2019;22(6):485–501.
8. Bullock S, Cliff D. Complexity and emergent behaviour in ICT systems. Technical Report HP-2004-187, Hewlett-Packard Labs; 2004. [This report was commissioned by the Foresight Programme of the UK's Office of Science and Technology (DTi). However, its findings are independent of government and do not constitute government policy].
9. Ladyman J, Lambert J, Wiesner K. What is a complex system? *Eur J Philos Sci*. 2013;3(1):33–67.
10. Lloyd S. Measures of complexity: a nonexhaustive list. *IEEE Contr Syst Mag*. 2001;21(4):7–8.
11. Mayfield M, Punzo G, Beasley R, Clarke G, Holt N, Jobbins S. Challenges of complexity and resilience in complex engineering systems. ENCORE Network+ White Paper. 2018.
12. Pariès J. Complexity, emergence, resilience.… In: Hollnagel, E, Woods, dd, Leveson, N, eds. *Resilience Engineering*. Farnham, UK: CRC Press; 2017:43–53.
13. Suh NP. Complexity in engineering. *CIRP Ann-Manuf Techn*. 2005;54(2):46–63.
14. Potts M, Sartor P, Johnson A, Bullock S. Through a glass, darkly? Taking a network perspective on system-of-systems architectures. In: Bonjour E, Krob D, Palladino L, Stephan F, eds. *Complex Systems Design & Management*. Paris, France: Springer International Publishing; 2019:121–132.
15. Broniatowski DA. Building the tower without climbing it: progress in engineering systems. *Syst Eng*. 2018;21(3):259–281. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/sys.21426.
16. Snowden DJ, Boone ME. A leader's framework for decision making. *Harv Bus Rev*. 2007;85(11):68.
17. Sargut G, McGrath RG. Learning to live with complexity. *Harv Bus Rev*. 2011;89(9):68–76.
18. Walden DD, Roedler GJ, Forsberg K, Hamelin RD, Shortell TM. *Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities*. Hoboken, New Jersey: Wiley; 2015.
19. Luzeaux D, Wippler JL. *Large-scale Complex System and Systems of Systems*. Hoboken, New Jersey: John Wiley & Sons; 2013.
20. Bar-Yam Y. *Dynamics of Complex Systems*. vol. 213. Reading, MA: Addison-Wesley; 1997.
21. Fischi J, Nilchiani R, Wade J. Dynamic complexity measures for use in complexity-based system design. *IEEE Syst J*. 2017;11(4):2018–2027.
22. Mostashari A, Sussman JM. A framework for analysis, design and management of complex large-scale interconnected open sociotechnological systems. *IJDSST*. 2009;1(2):53–68.
23. Sillitto H. *Architecting Systems: Concepts, Principles and Practice*. London, UK: College Publications; 2014.
24. Ameri F, Summers JD, Mocko GM, Porter M. Engineering design complexity: an investigation of methods and measures. *Res Eng Des*. 2008;19(2-3):161–179.
25. Sinha K, de Weck OL. Empirical validation of structural complexity metric and complexity management for engineering systems. *Syst Eng*. 2016;19(3):193–206.
26. Sinha K, de Weck OL. Structural complexity metric for engineered complex systems and its application. In: *Gain Competitive Advantage by Managing Complexity: Proceedings of the 14th International DSM Conference*. Kyoto, Japan. pp. 181–194.
27. Sinha K, Suh ES. Pareto-optimization of complex system architecture for structural complexity and modularity. *Res Eng Des*. 2018;29(1):123–141.
28. Sinha K, Suh ES, de Weck O. Integrative complexity: an alternative measure for system modularity. *J Mech Des*. 2018;140(5). https://doi.org/10.1115/1.4039119.
29. Broniatowski DA, Moses J. Measuring flexibility, descriptive complexity, and rework potential in generic system architectures. *Syst Eng*. 2016;19(3):207–221.
30. Lloyd S. Measures of complexity: a nonexhaustive list [Journal Article]. *IEEE Contr Syst Mag*. 2001;21(4):7–8.
31. Ladyman J, Lambert J, Wiesner K. What is a complex system? *Eur J Philos Sci*. 2013;3(1):33–67.
32. SEBoK. Complexity—SEBoK; 2018. [Online; accessed 24-January-2019]. Available from: https://www.sebokwiki.org/w/index.php?title=Complexity&oldid=54319.
33. Simpson JJ, Simpson MJ. System of systems complexity identification and control. In: *2009 IEEE International Conference on System of Systems Engineering (SoSE)*. IEEE; 2009. p. 1–6.
34. Simpson J, Simpson M. Complexity reduction: a pragmatic approach. *Syst Eng*. 2011;14(2):180–192.
35. Gilbert D, Yearworth M. Complexity in a systems engineering organization: an empirical case study. *Syst Eng*. 2016;19(5):422–435.
36. Barker BG, Verma D. Systems engineering effectiveness: a complexity point paradigm for software intensive systems in the information technology sector. *Eng Manag J*. 2003;15(3):29–35.
37. Ellinas C, Allan N, Johansson A. Toward project complexity evaluation: a structural perspective. *IEEE Syst J*. 2018;12(1):228–239.
38. Componation PJ, Youngblood AD, Utley DR, Farrington PA. A preliminary assessment of the relationships between project success, system engineering, and team organization. *Eng Manag J*. 2008;20(4):40–46.
39. Mirza E, Ehsan N. Quantification of project execution complexity and its effect on performance of infrastructure development projects. *Eng Manag J*. 2017;29(2):108–123.
40. Rouse WB. Complex engineered, organizational and natural systems. *Syst Eng*. 2007;10(3):260–71.
41. Tolk A, Diallo S, Mittal S. *Complex Systems Engineering and the Challenge of Emergence. Emergent Behavior in Complex Systems Engineering: A Modeling and Simulation Approach*. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2018:79–97. http://doi.org/10.1002/9781119378952.ch5
42. Martin JN. 3.1.2 The seven samurai of systems engineering: dealing with the complexity of 7 interrelated systems. *INCOSE Int Symp*. 2004;14(1):459–470.
43. Stevens R. Profiling complex systems. In: *2nd Annual IEEE Systems Conference*. IEEE; 2008:1–6.
44. Stacey R. *Tools and Techniques of Leadership and Management: Meeting the Challenge of Complexity*. Abingdon, UK: Routledge; 2012.
45. Beale D, Tryfonas T. Exploration of the complex ontology. In: *INCOSE International Symposium*. vol. 28. Wiley Online Library; 2018:1549–1563.
46. Beale D, Tryfonas T, Young M. Evaluating approaches for the next generation of difficulty and complexity assessment tools. In: *2017 IEEE Technology & Engineering Management Conference (TEMSCON)*. p. 227–233.
47. Summers JD, Shah JJ. Mechanical engineering design complexity metrics: size, coupling, and solvability. *J Mech Des*. 2010;132(2). http://doi.org/10.1115/1.4000759.
48. Alkan B, Vera DA, Ahmad M, Ahmad B, Harrison R. Complexity in manufacturing systems and its measures: a literature review. *Eur J Ind Eng*. 2018;12(1):116–150.
49. Min G, Suh ES, Hölttä-Otto K. System architecture, level of decomposition, and structural complexity: analysis and observations. *J Mech Des*. 2016;138(2). http://doi.org/10.1115/1.4032091.

50. Kim G, Kwon Y, Suh ES, Ahn J. Correlation between architectural complexity of engineering systems and actual system design effort. *J Mech Des*. 2017;139(3). http://doi.org/10.1115/1.4035319.
51. Boehm BW, Valerdi R. Achievements and challenges in cocomo-based software resource estimation. *IEEE Soft*. 2008;25(5):74–83.
52. Valerdi R. *The constructive systems engineering cost model (COSYSMO) [Thesis]*. University of Southern California; 2005.
53. McCabe TJ. A complexity measure. *IEEE Trans Softw Eng*. 1976;(4):308–320.
54. Kafura D, Reddy GR. The use of software complexity metrics in software maintenance. *IEEE Trans Softw Eng*. 1987;(3):335–343.
55. Allaire D, He Q, Deyst J, Willcox K. An information-theoretic metric of system complexity with application to engineering system design. *J Mech Des*. 2012;134(10). http://doi.org/10.1115/1.4007587.
56. Beale D, Tryfonas T. Assessing and developing complexity categorization frameworks. In: *2019 International Symposium on Systems Engineering (ISSE)*; 2019:1–8.
57. Stacey RD. *Complexity and Creativity in Organizations*. Oakland, California: Berrett-Koehler Publishers; 1996.
58. Thales Group. Complexity Profiler; 2013. Document Reference 87203605-DDQ-GRP-EN, Revision 003.
59. Potts M, Sartor P, Johnson A, Bullock S. Deriving key features of a system-of-systems complexity evaluation framework from an industrial case study analysis. In: *2019 International Symposium on Systems Engineering (ISSE)*; 2019:1–8.
60. Thales Group. Thales Complexity Profiler User Guide; 2015. Document Reference 87203444-DDQ-GRP-EN, Revision 004.
61. Valerdi R. *The Constructive Systems Engineering Cost Model (COSYSMO)*. University of Southern California; 2005.
62. Buede DM, Miller WD. *The Engineering Design of Systems: Models and Methods*. Hoboken, New Jersey: John Wiley & Sons; 2016.
63. Kossiakoff A, Sweet WN, Seymour SJ, Biemer SM. *Systems Engineering Principles and Practice*. vol. 83. Hoboken, New Jersey: John Wiley & Sons; 2011.
64. de Weck OL, Roos D, Magee CL, Vest CM. Appendix: Engineering Systems Terms and Definitions. In: *Engineering Systems: Meeting Human Needs in a Complex Technological World*. Cambridge, Massachusetts: MIT Press; 2011:185–193.
65. Bullock S, Potts M. 20190611-Survey results survey design engineer system complexity factors. The University of Bristol; 2019. Available from: https://doi.org/10.5523/bris.2mlbi0pc4rdiy250iavo7uh0om.
66. Kowalski C, INCOSE. Which aspects of the system's complexity are most significant for the project's success or failure? 2019. Available from: https://www.incose.org/events-and-news/incose-and-se-news/2019/04/15/which-aspects-of-the-system-s-complexity-are-most-significant-for-the-project-s-success-or-failure.
67. Nikolić D. Non-parametric detection of temporal order across pairwise measurements of time delays. *J Comput Neurosci*. 2007;22(1):5–19.
68. Saaty TL. Decision making–the analytic hierarchy and network processes (AHP/ANP). *J Syst Sci Syst Eng*. 2004;13(1):1–35.
69. Jamshidi M. System of systems engineering: new challenges for the 21st century. *Aero El Sys Mag*. 2008;23(5):4–19.
70. Gillespie T. *Systems Engineering for Ethical Autonomous Systems*. London, UK: Institution of Engineering and Technology; 2019.
71. Olechowski Alison L, Eppinger Steven D, Joglekar Nitin, Tomaschek Katharina. Technology readiness levels: Shortcomings and improvement opportunities. *Systems Engineering*. 2020;23(4):395–408.

AUTHOR BIOGRAPHIES

Matthew W. Potts is a PhD research student at the University of Bristol, United Kingdom, within the Aerospace Engineering Department. He received the BEng degree from the University of Southampton, United Kingdom in Electromechanical Engineering (2008–2011). He served in various roles in the Royal Air Force (2011–2016) as an engineering officer and has worked as an independent consultant. He is a registered incorporated engineer with the Institute of Engineering and Technology and is a student member of INCOSE.

Pia A. Sartor is senior lecturer in the Department of Aerospace Engineering, University of Bristol, United Kingdom. She received a BEng (Hons) degree in Aerospace Engineering from Ryerson University, Toronto, Canada (2006) and a PhD degree in Mechanical Engineering from the University of Sheffield, United Kingdom (2011). She worked in industry for six years, including as a senior systems engineer at Safran Landing Systems. She is a registered chartered engineer with the Institute of Mechanical Engineers.

Angus Johnson is the Thales Head of Systems Research and the industry lead on the Thales–Bristol Partnership in Hybrid Autonomous Systems Engineering. He has a background in radar spectrum sensing and multifunction RF systems engineering.

Seth Bullock is Toshiba Chair in Data Science and Simulation in the Department of Computer Science at the University of Bristol, United Kingdom. He holds a first degree in cognitive science (1993) and a doctorate in evolutionary simulation modeling (1997) from Sussex University, UK. His research expertise lies in complex systems simulation modeling. He has won over £28m in research and infrastructure funding in this area and has undertaken consultancy for the UK Government on complexity in ICT (2004) and financial systems (2012). He publishes in international peer-reviewed journals spanning health, social science, economics, biology, architecture, engineering, environmental science, computing, and physics. He was twice elected to the Board of Directors of the International Society for Artificial Life and serves on the editorial boards of the MIT Press journals *Adaptive Behavior*, and *Artificial Life*, and *Frontiers in Robotics & AI*. He has given invited keynote lectures in London, Paris, Athens, Melbourne, Madrid, Granada, and Tokyo.

How to cite this article: Potts MW, Sartor PA, Johnson A, Bullock S. Assaying the importance of system complexity for the systems engineering community. *Systems Engineering*. 2020;23:579–596. https://doi.org/10.1002/sys.21550

APPENDIX A: FLEISS' $\varkappa$

To calculate the degree of agreement between respondents on their ratings of complexity factor importance we use Fleiss' $\varkappa$. Fleiss' $\varkappa$ measures the degree of agreement in ratings beyond that which would be expected by chance.

Let N be the number of subjects to be rated ($N = 6$), n be the number of raters ($n = 55$), and k be the number of categories into which assignments are made ($k = 5$). The ratings are indexed by $i = 1, \ldots N$ and the categories are indexed by $j = 1, \ldots k$. Let n_{ij} represent the number of respondents who assigned the i-th subject to the j-th category. Fleiss' $\varkappa$ is given by:

$$\varkappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}, \tag{A.1}$$

where $1 - \bar{P}_e$ gives the degree of agreement that is attainable above chance, $\bar{P} - \bar{P}_e$ gives the degree of agreement actually achieved above chance. If respondents were in complete agreement then $\varkappa = 1$, and if there is no agreement among the respondents, other than what would be expected by chance, then $\varkappa \le 0$.

To calculate $\varkappa$, first calculate p_j, the proportion of all assignments which were to the j-th category:

$$p_j = \frac{1}{Nn}\sum_{i=1}^{N} n_{ij}. \tag{A.2}$$

Then calculate P_i, the extent to which raters agree for the i-th subject (i.e., compute how many rater–rater pairs are in agreement, relative to the number of all possible rater–rater pairs):

$$P_i = \frac{1}{n(n-1)}\left[\left(\sum_{j=1}^{k} n_{ij}^2\right) - (n)\right]. \tag{A.3}$$

Then, compute $\bar{P}$, the mean of the P_is, and $\bar{P}_e$ which go into Equation (A.1):

$$\bar{P} = \frac{1}{Nn(n-1)}\left(\sum_{i=1}^{N}\sum_{j=1}^{k} n_{ij}^2 - Nn\right), \tag{A.4}$$

$$\bar{P}_e = \sum_{j=1}^{k} p_j^2. \tag{A.5}$$

TABLE B.1 Modified Saaty Scale[68] used to aggregate pair-wise comparisons

Value	Definition
1	Equal importance in a pair. Corresponding to "They are equally unimportant" and "They are equally important"
3	Moderate Importance. Corresponding to "A is slightly more important than B"
5	Strong Importance. Corresponding to "A is much more important than B"
$\frac{1}{3}$	Corresponding to "B is slightly more important than A'
$\frac{1}{5}$	Corresponding to "B is much more important than A"

APPENDIX B: AGGREGATING PAIR-WISE COMPARISON RESPONSES

The pair-wise responses were converted to integer values using a modified Saaty scale, Table B.1 and stored as a matrix of pair-wise elements, C_ij.

A normalized pair-wise response matrix, X_ij, is then created by dividing each element of C_ij by the sum of the values in each column of C_ij.

$$X_ij = \frac{C_ij}{\sum_{i=1}^{n} C_ij}. \tag{B.1}$$

Row totals of X_ij are then summed and divided by the number of complexity factors evaluated (6) to generate a weighted 'priority vector" for each respondent, W_ij.

$$W_ij = \frac{\sum_{j=1}^{n} X_ij}{n}. \tag{B.2}$$

The weighted priority vector is averaged over all respondents and normalized between zero and one to obtain an overall ranking of complexity factor importance for the sample population.